

Evaluating Text Similarity Techniques for Matching Personal Employee Objectives

Mohab Ghanem, 08.06.2020 – Kick-off Guided Research

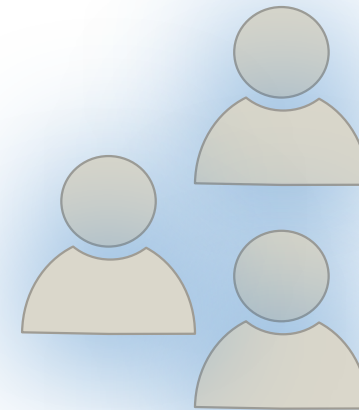
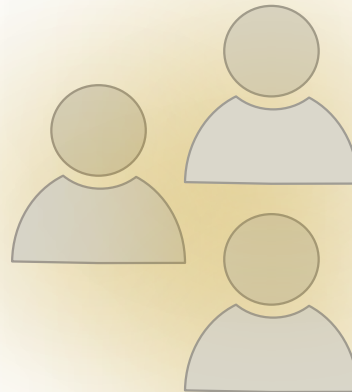
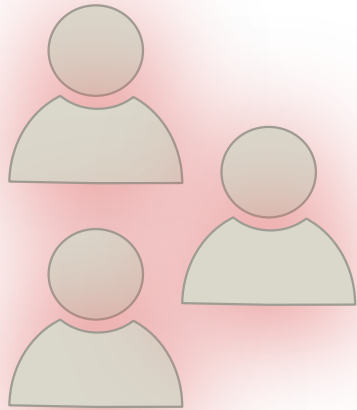
Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
www.matthes.in.tum.de

- 1) Problem Statement
- 2) Expected Output
- 3) Proposed Approach
- 4) Evaluating Results
- 5) Preliminary Results
- 6) Timeline
- 7) Questions

Problem Statement

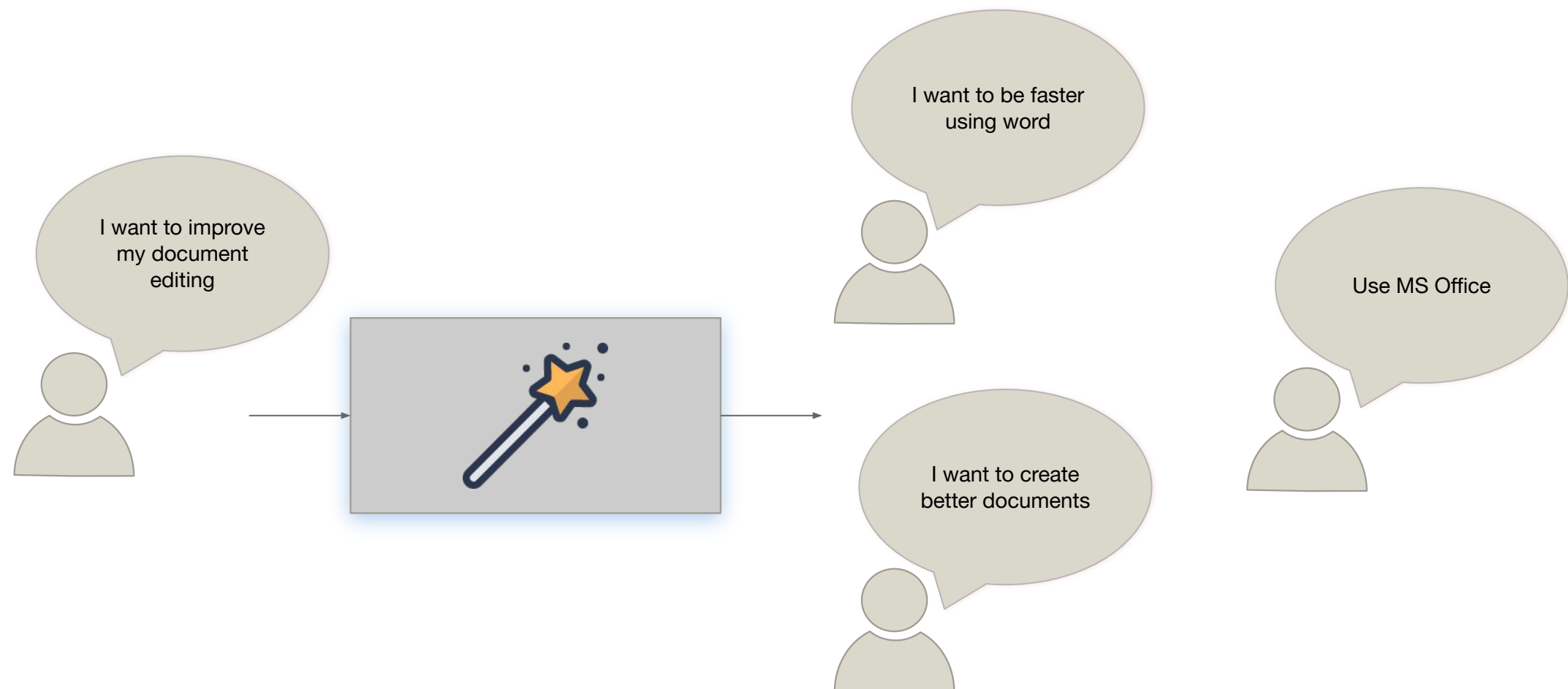
Employees find it difficult to achieve their personal objectives alone.

Text Similarity can match employees with similar goals.



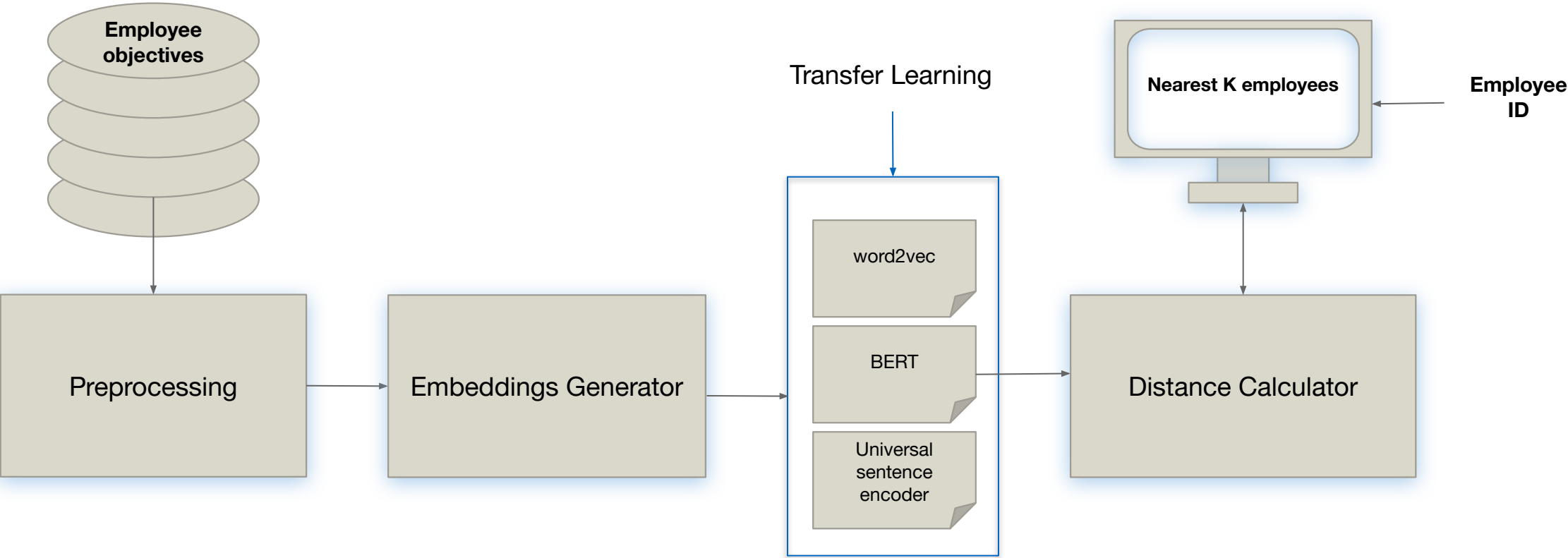
Expected Output

Given the employee's objective, find the most similar K employees.

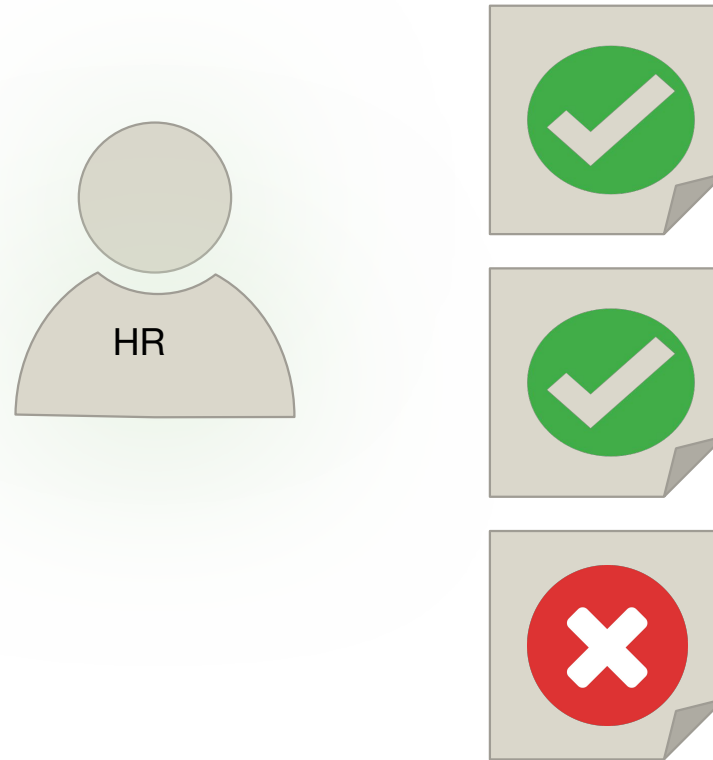


Proposed Approach

A three-part pipeline, from raw data to front end user interface



1. The software is to be used and manually evaluated by the HR department of MERCK



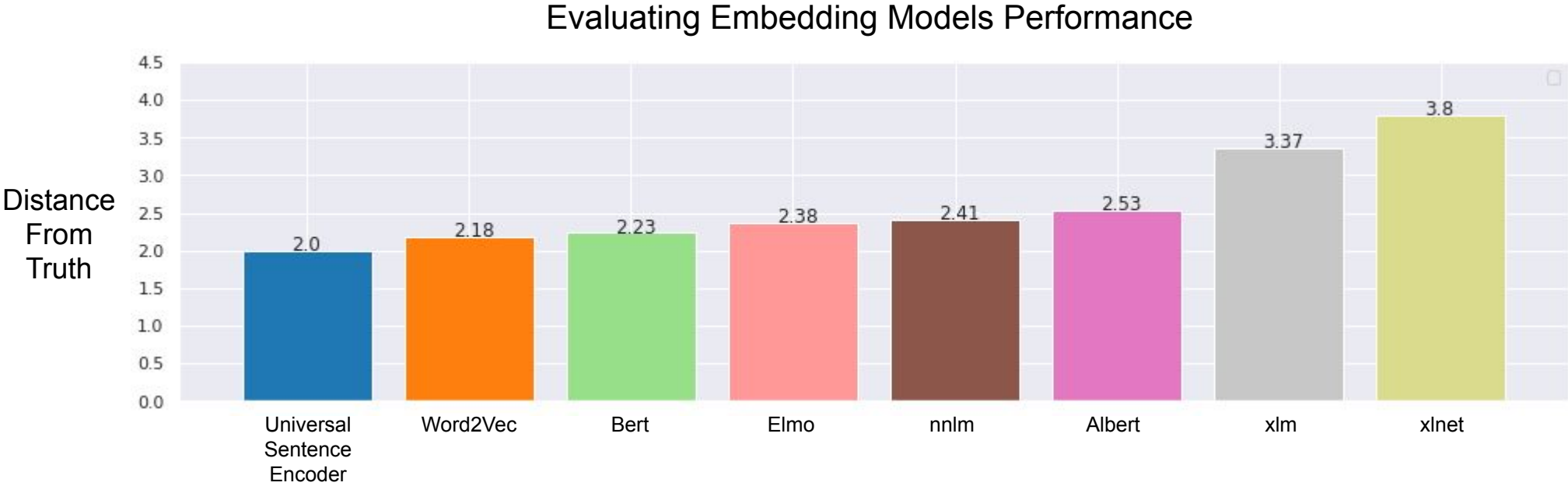
2. Using open-source annotated text similarity corpus.

Sentence 1	Sentence 2	Score
the young boys are playing outdoors and the man is smiling nearby	the kids are playing outdoors near a man with a smile	4.7
...	...	...
two young women are sparring in a kickboxing fight	two women are sparring in a kickboxing match	4.9

* 20K annotated pairs.

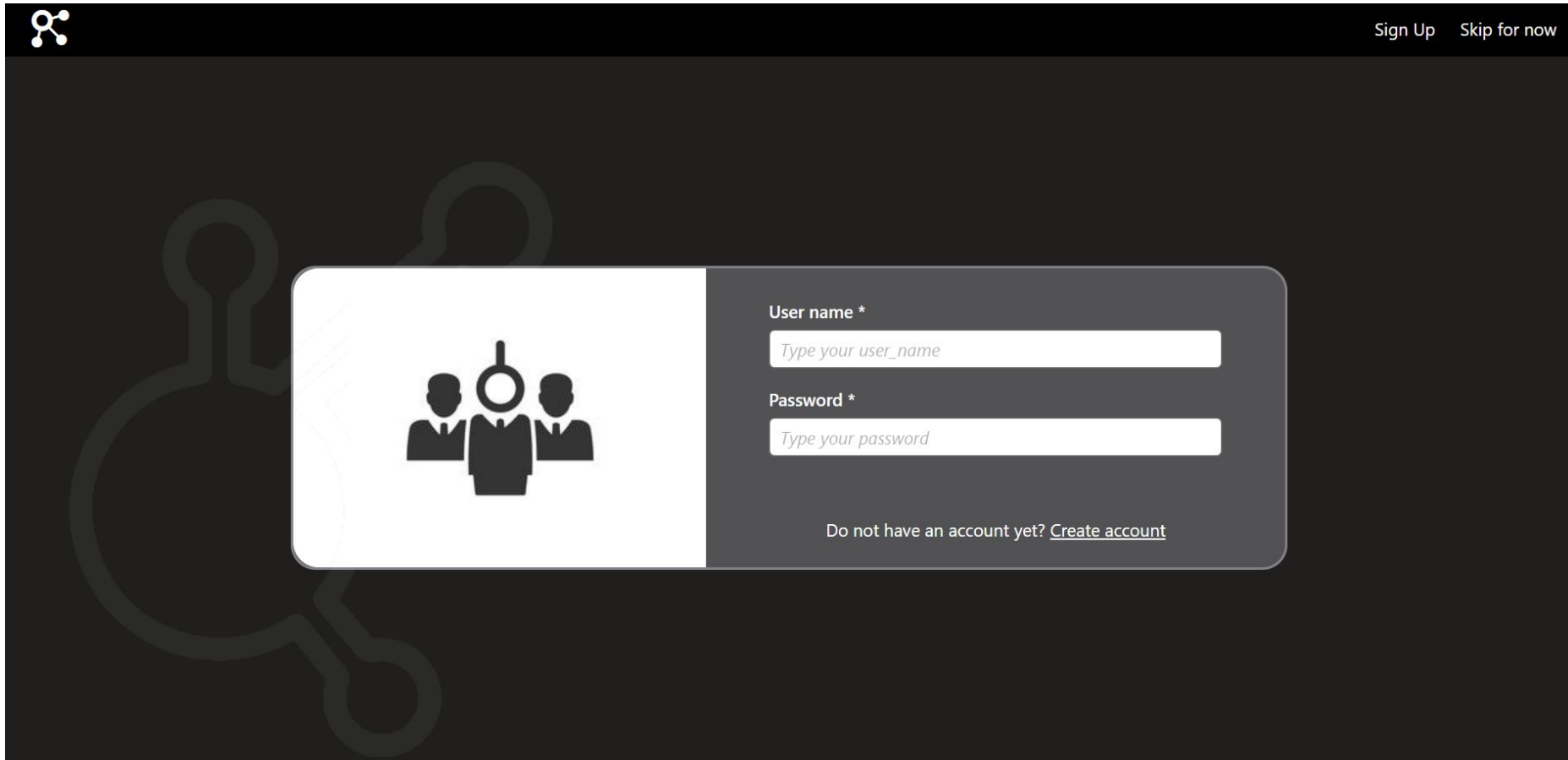
* <https://github.com/PrincetonML/SIF/blob/master/data/sicktest>

1. Analytical Results :



Preliminary Results

2. User Interface:



The image shows a user interface mockup for a login or registration page. The background is dark gray with a faint, large watermark of a person icon. In the top left corner, there is a small white icon of three people. In the top right corner, there are two links: "Sign Up" and "Skip for now". The main content area features a white rounded rectangle on the left containing an icon of three people with a magnifying glass over the center person. To the right of this is a dark gray rounded rectangle containing two input fields. The first field is labeled "User name *" and has a placeholder text "Type your user_name". The second field is labeled "Password *" and has a placeholder text "Type your password". Below these fields, there is a link that says "Do not have an account yet? [Create account](#)".

2. User Interface:

Embeddings Generation Options

Embeddings

Metrics

Upload data file

☒ use

☐ word2vec

☒ nnlm

☐ bert

☐ albert

☒ elmo

☐ xlm

☐ xlnet

Embeddings Generation Options

Embeddings

Metrics

Upload data file

☒ cosine

☐ minkowski

☐ correlation

☒ jensenshannon

☐ cityblock

☐ sqeuclidean

☐ euclidean

☒ chebyshev

☐ canberra

☐ braycurtis

Embeddings Generation Options

Embeddings

Metrics

Upload data file

File *

Choose File

No file chosen

ID field name

ID

Sentence feild name

Objective

Max K

10

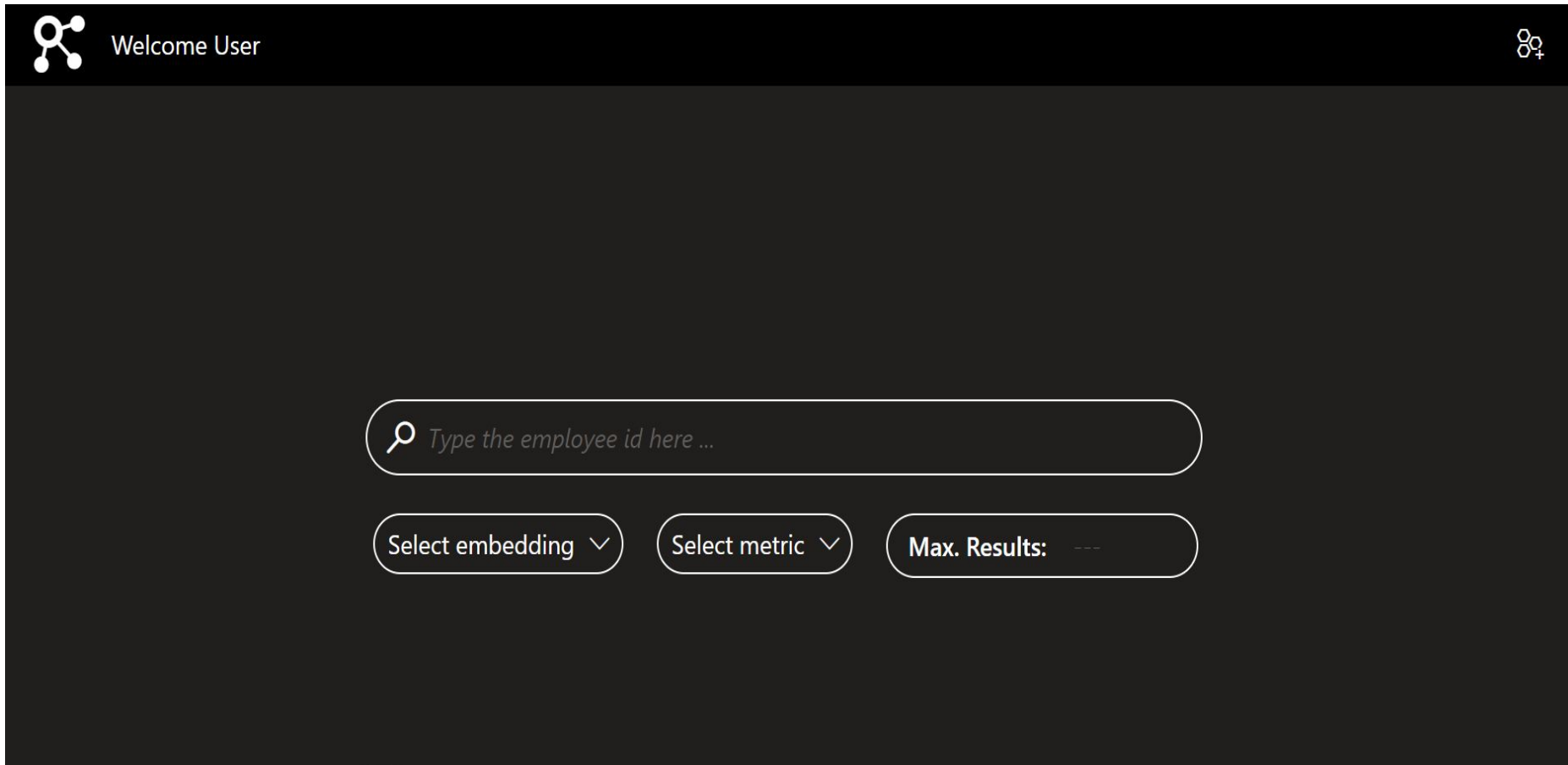
Chunk size

20

Generate

Preliminary Results

2. User Interface:



The image shows a dark-themed user interface for a search application. At the top left, there is a logo consisting of three connected dots and the text "Welcome User". At the top right, there is a small icon of three connected dots. In the center of the interface, there is a large, rounded rectangular search bar with a magnifying glass icon on the left and the placeholder text "Type the employee id here ...". Below the search bar, there are three rounded rectangular buttons. The first button is labeled "Select embedding" with a downward arrow. The second button is labeled "Select metric" with a downward arrow. The third button is labeled "Max. Results:" followed by three dashes.

2. User Interface:

Welcome User

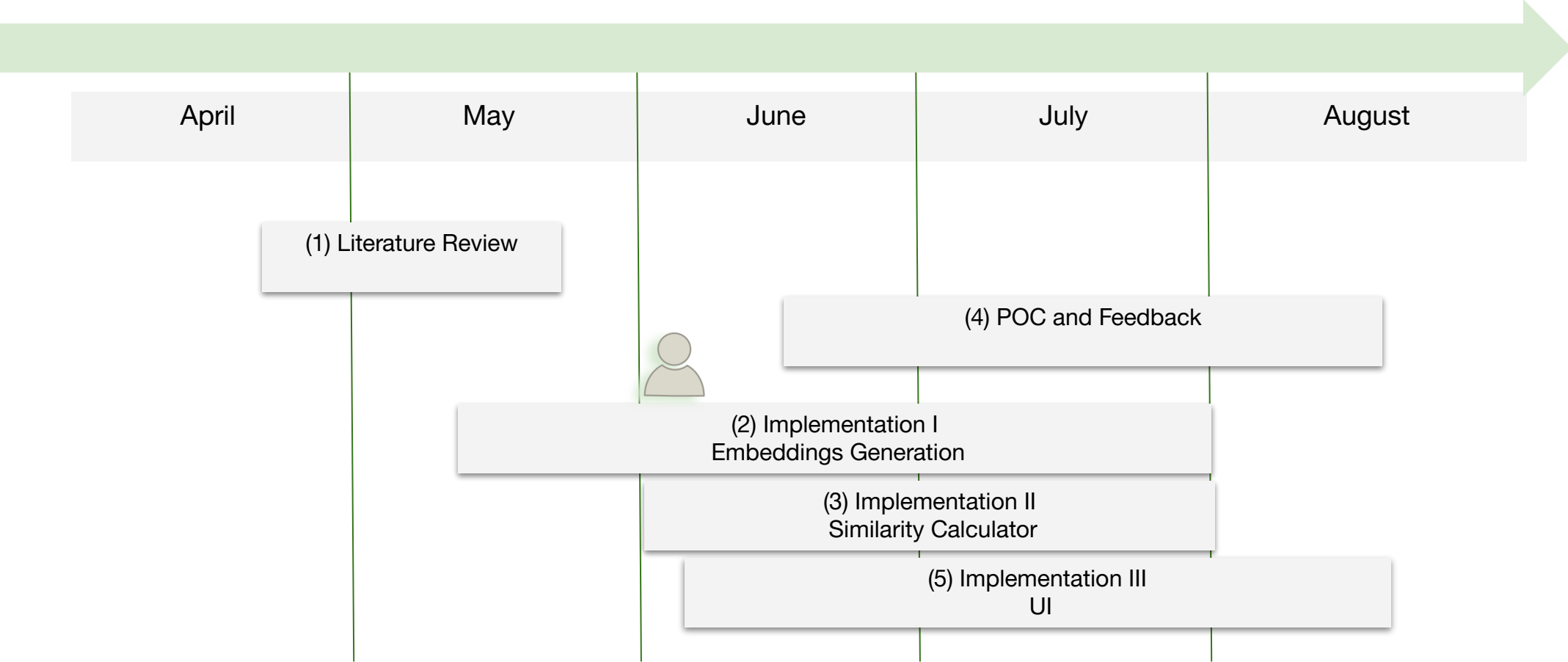
 a woman is peeling a potato 

use 

cosine 

Max. Results: 7

Employee id	Distance	Objective
4	0.6063145697	the pepper is being sliced by a woman with a knife
7	0.6522007585	a woman is penciling on eyeshadow
9	0.6543130279	a frog is being handled by the woman
2	0.7514517605	two women are wearing bikinis and walking on the sand
1	0.796306029	two women with folded arms , who are young , are looking at the screen
6	0.8036073744	makeup is being put on the man by the woman
3	0.8671488166	two children are crouching under some bars made of metal



Questions

Thank you!



Mohab Ghanem

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business
Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289. 17132
Fax +49.89.289.17136

matthes@in.tum.de
www.matthes.in.tum.de

